Data Analysis

Tong Zou

1 Abstract

Retest effect potentially obscures diagnoses on patients' cognitive performance through repeated measurement from neuropsychological tests. In this study we aim to quantify such effects on three cognitive tests (LMT, TMT, BNT) based on a large cohort longitudinal dataset across nation. Generalized linear model is used to analyze the data. Result shows that in first two visits, LMT scores increased significantly for both normal group (0.43, p=.00) and mild impaired group (0.47, p=.00). Also BNT scores increased significantly for both normal group (0.20, p=.00) and mildly impaired group (0.11, p=.02). In longitudinal analysis, rate of change in LMT scores increased significantly for both normal group (0.24/year, p=.00) and mildly impaired group (0.26/year, p=.00). Demented group doesn't show retest effect on three tests.

2 Introduction

Diagnoses of Alzheimer disease (AD) on the living patients usually require longitudinal cognitive tests to establish evidence of cognitive impairment. Nevertheless, repeated testing might yield in improved subsequent performances, known as retest effect or practice effect, confounding the major scientific interest in cognitive impairment.

Using alternative test forms may attenuate retest effects. Typical cognitive tests include Wechsler Memory Scale (WMS) designed to measure different memory functions, of which Logical Memory test is a subset with score ranging from 0 to 25. Study shows that Logical Memory Test (LMT) is consistently a reliable (less retest effect) subtest across the age cohorts and largest practice effects are found in the youngest age cohort [1]. The Trail Making Test (TMT) is a neuropsychological test ranging up to 300, providing information about visual search speed, scanning, speed of processing, mental flexibility, as well as executive functioning [2]. A linear growth analysis yielded statistically significant average change in slope across time periods for Trail Making test [3]. The Boston Naming Test (BNT) is widely used to measure confrontational word retrieval in individuals with aphasia or other language disturbance caused by stroke, Alzheimer's disease, or other dementing disorder [4]. A clinical study revealed high stability of the BNT scores over time, which suggests a lack of the practice effect [5]. Previous study suggests that alternate forms may attenuate but do not eliminate practice and retest effects [6].

Another approach is to quantify retest effects for typical neuropsychological tests to improve clarity in diagnoses. This study aims to use data from a large nation cohort study to quantify retest effects, accounting for different cognitive status. Main factors that may be connected with AD hence test performance include race, education, age [1], heavy smoking habit and alcohol use [7] [8], diabetes [9], hypertension [10], heart disease [11], seizures [12], etc. In order to identify retest effects from other possible factors, adjustment of these factors are necessary.

3 Method

3.1 Source of the Data

Longitudinal data was collected from different study centers across nation on 64864 observations of 15,665 unique subjects, each with three cognitive test performance (LMT, TMT, BNT). For those tests, questions with each test do not change from one visit to another. Tests were scheduled approximately once a year, with variation in the times at which subjects returned for visits. Number of visits and days of interval between them were recored. Subjects were characterized with normal, mild cognitive impairment (MCI) and AD as their cognitive status (may be time-varying).

Other available covariates are demographics including age, gender, race, education, personal habits (smoking and alcohol), co-morbidities including diabetes, stroke, seizures, trauma, arthritis, heart disease.

Covariates concerning alcohol, arthritis and urinary incontinence are removed due to critical missing. Proportion of missing observations in other covariates of interest are all below 5% compared to total observations: race (342, .5%), education years (233, .4%), smoked packs per day (1726, 2.7%), history of diabetes (197, .3%), hypertension (161, .2%), high cholesterol (688, 1%), traumatic brain injury (492, .8%). Other co-morbidities are not selected due to their extremely imbalanced distribution. Removal of those missingness (1939 observations, 489 subjects, 3%) will not have much impact on result of analysis due to their small proportions, although the pattern of missing may be non-ignorable (depending on missing value). Therefore data analysis is conducted as a complete fashion.

3.2 Statistical Models

To investigate the causal effect of retesting on test scores with longitudinal data, I use generalized linear model for dependent data, to acquire robust variance estimator of regression parameters, therefore more reliable inference results. In this study GEE package in R is used to achieve this object. To apply GEE model, both assumptions of mean model and covariance structure are required. Choice of mean model will be addressed later in this section, now denoted as $E(\vec{Y_i}) = X_i\vec{\beta}$ or $E(Y_{ij}) = \sum_k^p X_{ijk}\beta_k$, where i, j, k are indexes of subject, visit and parameter respectively. GEE acquires the estimates by performing iterative weighted least square (IWLS) to solve the estimating equation $Q_W(\vec{\beta}) = \sum_i^N (\vec{Y_i} - X_i\vec{\beta})^T W_i (\vec{Y_i} - X_i\vec{\beta})$, where W_i is the covariance structure that will be specified in Result Section, and estimated using method of moment (MOM). In order to acquire valid inference, Huber-White robust variance estimator will be used to estimate the variance of regression parameter.

To investigate the short-term impact of retesting and on test score Y with respect to different cognitive status, my mean model for i-th subject is

$E(Y_{ij}) = \beta_0 + \beta_1 j + \beta_2 dx_{ij} + \beta_3 j dx_{ij} + others,\ j = 0, 1$

where j is the indicator of the second visit (0 = 1st visit) and dx_{ij} the factor variable of cognitive status (normal as reference, MCI, AD). Note that dx might be variant within subject, which would cause ambiguity when describing trajectories stratified by cognitive status, but it distinguishes the retest effect from cognitive impairment effect on test, and the interpretations of regression parameters do not change, hence necessary to adjust.

β_1 would be of scientific interest, denoting the effect of retesting on test performance for normal group. β_3 measures the difference in effect of retesting between normal and MCI (or AD) groups. The rest of the regression parameters are all adjusted to improve prediction precision, including age at baseline, gender, race (white as reference, African American and other), years of education, packs smoked per day (0 as reference, < 1, $\geqslant 1$), history of diabetes, hypertension, high cholesterol and traumatic brain injury (Absent as reference, Remote/Inactive regrouped as Absent).

To further quantify the rate of change in test scores using longitudinal data, my mean model for i-th subject is:

$E(Y_{ij}) = \beta_0 + \beta_1 year_{ij} + \beta_2 dx_{ij} + \beta_3 year_{ij} dx_{ij} + others,\ j = 0, 1, 2...$

where $year_{ij}$ is the number years between first visit (j=0) and current visit j. Here number of visit is not used for its collinearity with $year_{ij}$ (corr = 0.97), so change of rate in test score is defined as retest effect. β_1 then quantifies the rate of change in test scores per year for the normal group. β_3 measures the difference in rate of change in test scores between normal and MCI (or AD) groups. β_1 and β_3 jointly determine whether retest effect exist in different diagnosis groups.

For the first model an exchangeable covariance structure is applied due to only two observations with each subjects. Further exploratory data analyses are required to determine the covariance structure for the second model.

4 Results

4.1 Exploratory Data Analysis

Table 1 presents summaries of variables of interest for 15176 subjects, reflecting the sampling scheme of data. For continuous variables, medians and inter-quartile ranges are presented due to

the heavy skewness of variables such as age and education years. Other variables are abandoned due to critical missingness (alcohol, arthritics and urinary incontinence) or extremely imbalanced distribution (stroke, seizure, heart disease). Therefore data generalizes to the white, healthy, well-educated elder populations. Table 2 presents the number of subjects with a given observations-per-subject. Note that number of subjects rapidly decreases as number of measurement grows.

Figure 1, 2, 3 and 4 are respectively mean plots (across all subjects) and spaghetti plots (750 randomly sampled subjects) of three test scores for different treatment groups. Here I use all observations to describe the averaged effect of three groups. As for spaghetti plots, which describe within-subject trajectories, only subjects with consistent longitudinal group attribute are sampled.

Mean plots of test scores show significant difference among three diagnosis group. For LMT and BNT, improved test performances can be observed for all groups, but not noticeable for TMT. Spaghetti plots can only differentiate three groups in a general sense. The change of test performance is not very obvious. Only an increasing trend in TMT score for AD group is noticeable, which might be accounted by aging effect on cognitive performance.

In the following I will specify the covariance structure. Figure 5, 6 and 7 show pairwise residuals plots of three test scores over time. Correlation exists between neighboring years, and becomes slightly weaker as time difference grows. Plot of TMT shows some non-linear correlation.

Variograms 8 shows that there are noticeable serial processes in the covariance structure, and also components of random intercept and measurement error cannot be ignored. Hence I will choose auto-regression (AR) as the covariance structure for longitudinal data, which specifies constant measurement error and correlation and exponentially decayed serial processes between measurements across all subjects.

4.2 Statistical Inference

Below are regression parameters estimates from generalized linear models addressed in Statistical Method Section, along with summary on 95% confidence interval and P-value.

- Short-term (two visits): Results are show in table 3, 4 and 5. For LMT, retest effect is significant for normal group (0.43, p=.00). Linear contrast method is used to determine the retest effect for MCI and AD groups since their retest effects differ significantly from normal group. For MCI and AD group, retest effects are respectively 0.47 (p=.00) and -0.51 (p=.00).

  For TMT, retest effect is not significant for normal group (-0.22, p=.70). AD group differs significantly from normal group (19.26, p=.00), but not for MCI group (1.96, p=.07). Results of linear contrast show that retest effects of MCI and AD groups are respectively 1.74 (p=.08) and 19.04 (p=.00).

  For BNT, retest effect is significant for normal group (0.20, p=.00). AD group differs significantly from normal group (-1.33, p=.00), but not for MCI group (-0.09, p=.13). Results of linear contrast show that retest effects of MCI and AD groups are respectively 0.11 (p=.02) and -1.13 (p=.00).

- Longitudinal: Results are show in table 6, 7 and 8. In the longitudinal context, change rate of test score can be seen as retest effect. For LMT, retest effect is significant for normal group (0.24, p=.00). Retest effect of AD group differs significantly from normal group (-0.27, p=.00). Linear contrast method is used to determine the retest effect for MCI and AD groups. For MCI and AD group, retest effects are respectively 0.26 (p=.00) and -0.03 (p=.09).

  For TMT, retest effect is not significant (TMT score should decrease as a sign of retest effect) for normal group (1.15, p=.00). Both MCI and AD group differs significantly from normal group. Results of linear contrast show that retest effects of MCI and AD groups are respectively 2.56 (p=.00) and 6.37 (p=.00).

  For BNT, retest effect is not significant for normal group (0.01, p=.16). Both MCI and AD group differs significantly from normal group. Results of linear contrast show that retest effects of MCI and AD groups are respectively -0.05 (p=.00) and -0.44 (p=.00).

Results of retest effect are summarized in table 9 and 10.

5 Discussion

According to table 9 and 10, for short-term analysis, normal and MCI groups have consistent retest effect on LMT and BNT. AD group shows declined performance for all three tests. For longitudinal analysis, all three groups have increased rate of change (declined performance) in TMT. Normal and MCI groups have increased rate of change (retest effect) in LMT. MCI and AD groups have decreased rate of change in BNT.

To sum up, LMT has strongest retest effect compared to TMT (weakest) and BNT (moderate). Normal and MCI groups have stronger retest effect than AD group. Short-term test have stronger retest effect than long-term test. The final results seem to disagree with mean plot, especially in the tail area. The decreasing trend is obvious at starting point for AD group. This arises from small number of observations for large number of visits, hence smaller regression weight. In such sense, analysis on the short-term data are more reliable. Based on the analysis, LMT and BNT should be adjusted for retest effect for subjects in normal and MCI groups, with different questions at each visit, or penalized scores for repeated visit.

In terms of limitation of this study, aside from lack of follow-up observations, the most important fact is we are dealing with observational data with potential confounding. For example, aging is a factor affecting test performance which may cover up retest effect, thus adjusting for aging is necessary to identify retest effect. However aging is highly correlated with retesting, resulting in variance inflation for inference, thus difficult to separate two effects. One way to reduce aging effect is to recruit younger subjects, which is another limitation of the data set. Also the diagnosis groups are likely defined by test score, which might be mixed with retest effect already, so using such diagnosis result would introduce more bias.

To further investigate the retest effect more accurately, researchers should consider randomized controlled study with more complete follow-ups to reduce other confounding effect associated with retesting. Younger subjects and shorter interval between visits should be taken into consideration to separate aging effect.

6 Appendix

Variables	Median (IQR) or N (%)
age	72.45 (66.00, 79.00)
female	8713 (57.4%)
race (white)	12540 (82.63%)
race (AA)	1792 (11.81%)
race (other)	844 (5.56%)
education years	16 (13, 18)
packs smoked per day	
(= 0)	8049 (53.04%)
(< 1)	2218 (14.62%)
(⩾ 1)	4909 (32.35%)
diabetes	1741 (11.47%)
hypertension	7389 (48.69%)
high cholesterol	7509 (49.48%)
trauma	1578 (10.40%)
baseline diagnosis	
(normal)	9082 (59.84%)
(MCI)	3369 (22.20%)
(AD)	2725 (17.96%)

Table 1: Variables summary

Number of measurement	2	3	4	5	6	7	8	9	10	11	12
Number of subjects	4595	3202	2211	1431	1193	908	698	529	257	123	29

Table 2: Number of subjects with a given observations-per-subject.

7 Reference

[1] Lo N.A., Humphreys A.H., Byrne M., Pachana G.J. Test-retest reliability and practice effects of the Wechsler memory scale-III. J. Neuropsychol. 2012;6:212–231.

[2] Arnett, James A.; Seth S. Labovitz (1995). Effect of physical layout in performance of the Trail Making Test. Psychological Assessment. 7 (2): 220–221. doi:10.1037/1040-3590.7.2.220. Retrieved 2012-02-22.

[3] Buck, Kristen K., Thomas M. Atkinson, and Jeanne P. Ryan. Evidence of practice effects in variants of the Trail Making Test during serial assessment. Journal of clinical and experimental neuropsychology 30.3 (2008): 312-318.

[4] Kaplan; Goodglass, Harold; Weintraub, Sandra (1983). Boston Naming Test. Philadelphia: Lea & Febiger. OCLC 10450471.

[5] Mitrushina, M. & Satz, P. Aging Clin Exp Res (1995) 7: 123.

[6] Beglinger L.J., Gaydos B., Tangphao-Daniels O., Duff K., Kareken D.A., Crawford J. Practice effects and the use of alternate forms in serial neuropsychological testing. Arch Clin Neuropsychol. 2005;20:517–529.

[7] Durazzo, Timothy C., Niklas Mattsson, and Michael W. Weiner. Smoking and Increased Alzheimer's Disease Risk: A Review of Potential Mechanisms. Alzheimer's & dementia: the journal of the Alzheimer's Association 10.3 0 (2014): S122-S145. PMC. Web. 29 June 2018. [8] Tyas, Suzanne L. Alcohol use and the risk of developing Alzheimer's disease. Alcohol Research and Health 25.4 (2001): 299-307.

[9] Ojo O, et al. Evaluating the association between diabetes, cognitive decline and dementia. International Journal of Environmental Research and Public Health. 2015;12:8281.

[10] Kivipelto, Miia, et al. Hypertension and hypercholesterolaemia as risk factors for Alzheimer's disease. CNS drugs 16.7 (2002): 435-444.

[11] Cermakova, P., et al. Heart failure and Alzheimer' s disease. Journal of internal medicine 277.4 (2015): 406-425.

[12] Born, H. A. Seizures in Alzheimer's disease. Neuroscience 286 (2015): 251-263.

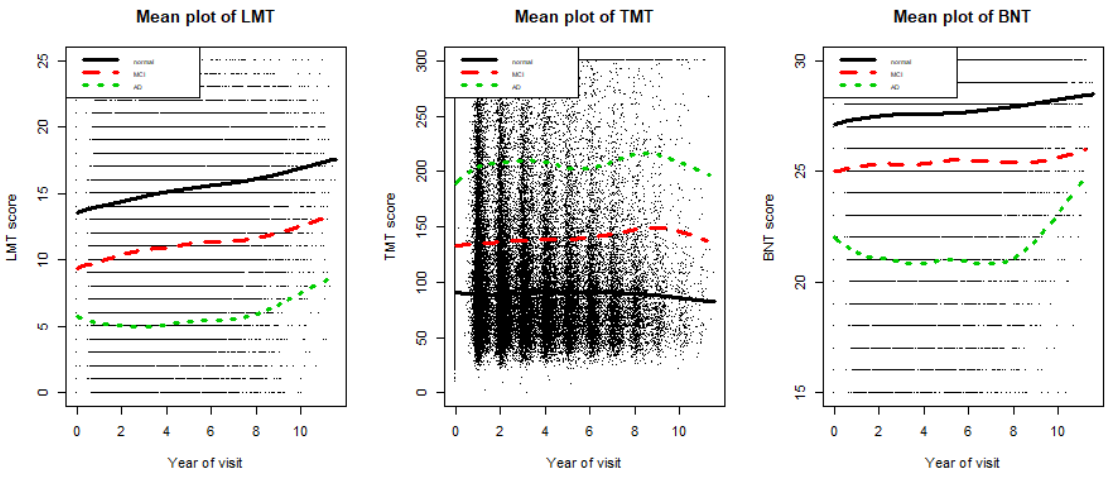

Figure 1: Meanplot of tests scores stratified by diagnosis groups

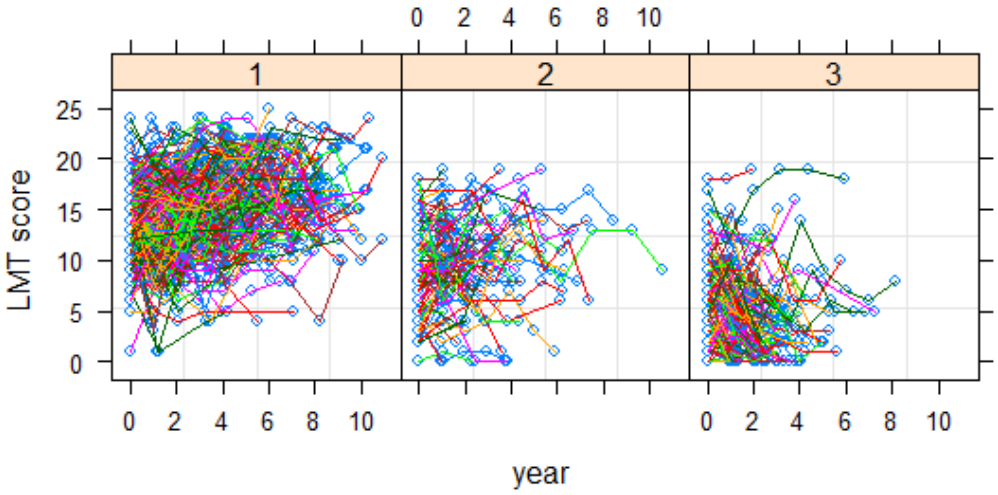

Figure 2: Spaghetti plot of LMT score stratified by diagnosis groups

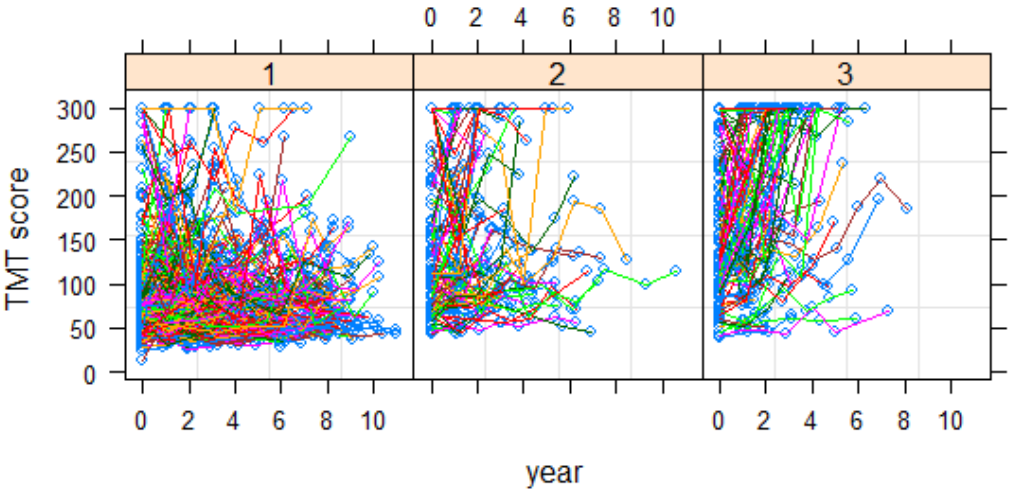

Figure 3: Spaghetti plot of TMT score stratified by diagnosis groups

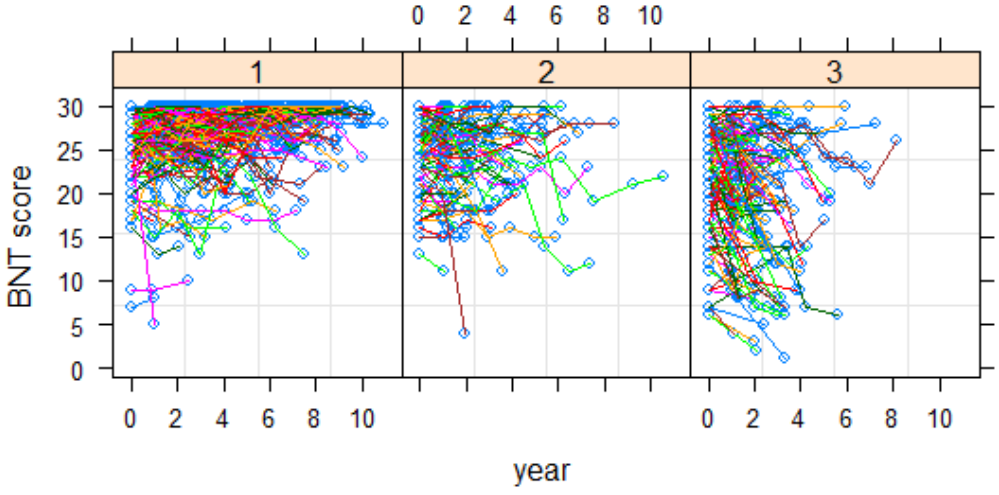

Figure 4: Spaghetti plot of BNT score stratified by diagnosis groups

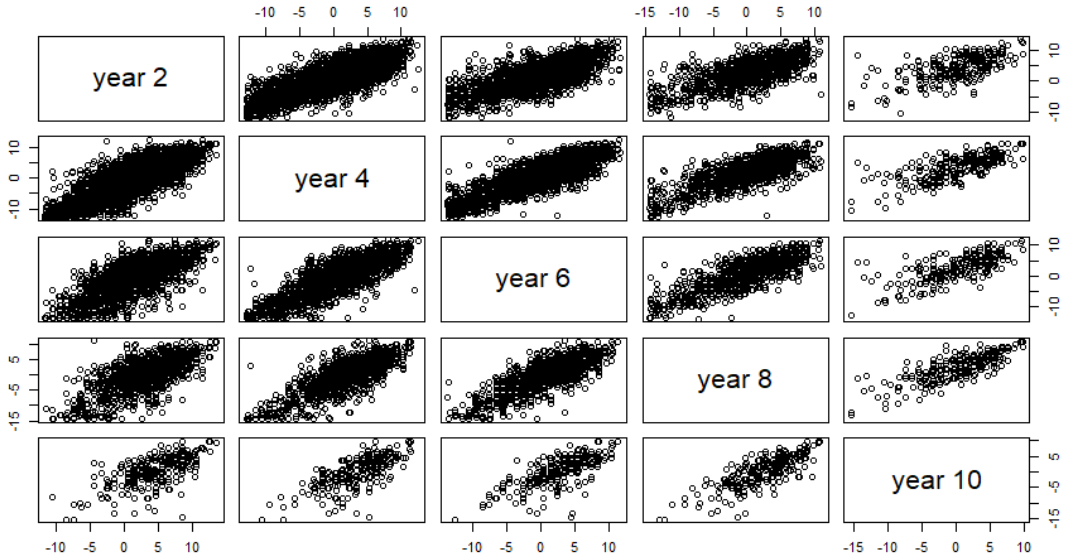

Figure 5: Pairwise residuals of LMT score over time

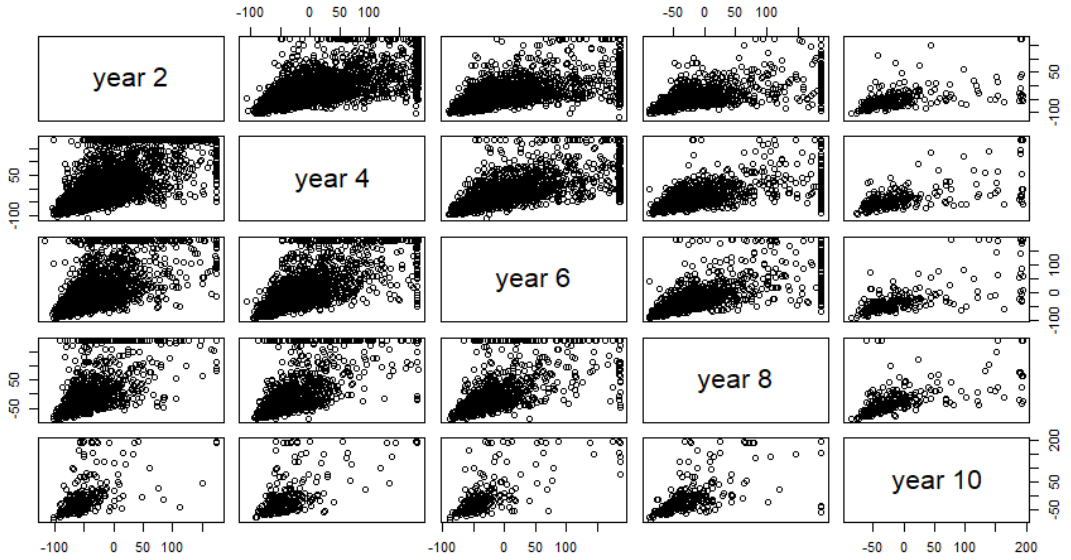

Figure 6: Pairwise residuals of TMT score over time

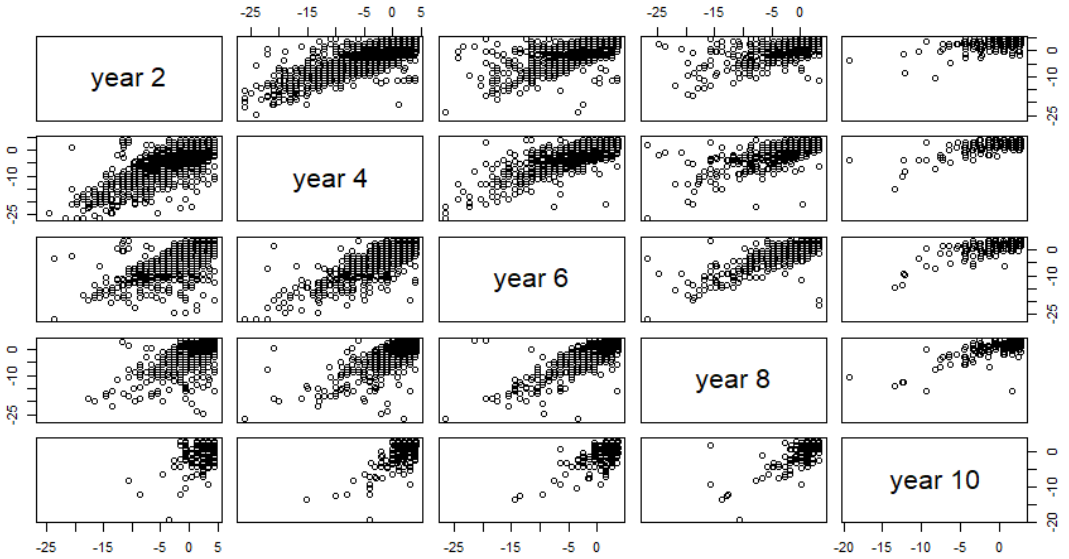

Figure 7: Pairwise residuals of BNT score over time

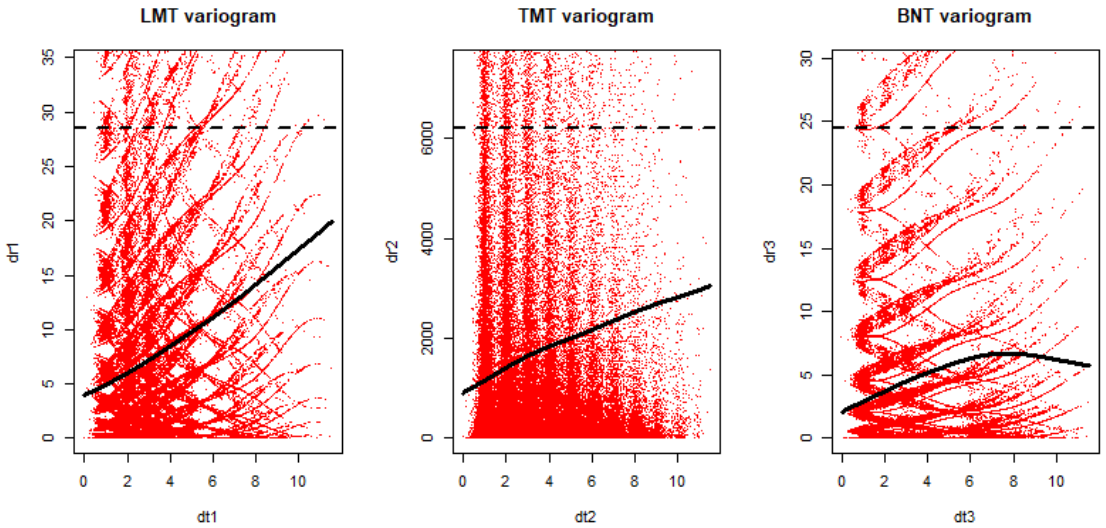

Figure 8: Variograms of three test scores over time

	Est	ci95.lo	ci95.hi	z value	Pr(>|z|)
(Intercept)	13.01	12.47	13.55	46.84	0.00
j	0.43	0.36	0.51	11.10	0.00
dx(MCI)	-3.40	-3.53	-3.27	-50.61	0.00
dx(AD)	-6.59	-6.74	-6.43	-84.37	0.00
j*dx(MCI)	0.04	-0.10	0.18	0.56	0.57
j*dx(AD)	-0.94	-1.09	-0.80	-12.62	0.00
age	-0.06	-0.07	-0.06	-20.92	0.00
female	1.03	0.91	1.15	17.42	0.00
race(AA)	-1.34	-1.52	-1.16	-14.82	0.00
race(other)	-0.87	-1.11	-0.63	-7.12	0.00
education years	0.26	0.24	0.28	27.68	0.00
pack(<1)	0.24	0.08	0.40	2.98	0.00
pack($\geqslant$ 1)	0.35	0.22	0.47	5.54	0.00
diabetes	0.18	-0.00	0.35	1.94	0.05
hypertension	0.00	-0.12	0.12	0.04	0.97
high cholesterol	-0.05	-0.17	0.06	-0.90	0.37
trauma	0.08	-0.09	0.26	0.92	0.36

Table 3: Regression estimates on LMT score over first two visits

	Est	ci95.lo	ci95.hi	z value	Pr(>|z|)
(Intercept)	36.76	27.97	45.55	8.20	0.00
j	-0.22	-1.37	0.92	-0.38	0.70
dx(MCI)	35.64	33.60	37.67	34.33	0.00
dx(AD)	80.64	78.25	83.04	65.99	0.00
j*dx(MCI)	1.96	-0.18	4.09	1.80	0.07
j*dx(AD)	19.26	17.08	21.45	17.28	0.00
age	1.84	1.75	1.93	38.61	0.00
female	-1.52	-3.39	0.36	-1.59	0.11
race(AA)	38.31	35.45	41.16	26.27	0.00
race(other)	16.13	12.26	19.99	8.18	0.00
education years	-5.05	-5.34	-4.75	-33.17	0.00
pack(<1)	0.47	-2.12	3.06	0.36	0.72
pack($\geqslant$ 1)	-3.80	-5.78	-1.81	-3.75	0.00
diabetes	7.65	4.77	10.53	5.21	0.00
hypertension	2.23	0.33	4.14	2.30	0.02
high cholesterol	-2.25	-4.09	-0.41	-2.40	0.02
trauma	-2.18	-5.07	0.71	-1.48	0.14

Table 4: Regression estimates on TMT score over first two visits

	Est	ci95.lo	ci95.hi	z value	Pr(>|z|)
(Intercept)	29.61	28.98	30.24	92.15	0.00
j	0.20	0.14	0.26	6.34	0.00
dx(MCI)	-1.64	-1.76	-1.51	-25.95	0.00
dx(AD)	-3.58	-3.73	-3.43	-45.93	0.00
j*dx(MCI)	-0.09	-0.21	0.03	-1.53	0.13
j*dx(AD)	-1.33	-1.44	-1.21	-22.01	0.00
age	-0.09	-0.10	-0.08	-26.35	0.00
female	-0.57	-0.70	-0.43	-8.31	0.00
race(AA)	-2.99	-3.20	-2.79	-28.62	0.00
race(other)	-1.90	-2.18	-1.63	-13.45	0.00
education years	0.27	0.25	0.29	24.54	0.00
pack(<1)	0.12	-0.06	0.31	1.30	0.19
pack($\geqslant$ 1)	0.39	0.25	0.53	5.37	0.00
diabetes	0.01	-0.19	0.22	0.12	0.90
hypertension	0.11	-0.02	0.25	1.62	0.10
high cholesterol	0.06	-0.07	0.19	0.91	0.37
trauma	0.32	0.11	0.52	2.98	0.00

Table 5: Regression estimates on BNT score over first two visits

	Est	ci95.lo	ci95.hi	z value	Pr(>|z|)
(Intercept)	13.03	12.54	13.51	52.26	0.00
year	0.24	0.22	0.26	25.07	0.00
dx(MCI)	-2.87	-2.99	-2.76	-50.58	0.00
dx(AD)	-6.08	-6.22	-5.95	-87.28	0.00
year*dx(MCI)	0.02	-0.02	0.05	0.99	0.32
year*dx(AD)	-0.27	-0.31	-0.23	-13.43	0.00
age	-0.07	-0.08	-0.07	-27.41	0.00
female	1.25	1.15	1.36	24.01	0.00
race(AA)	-1.37	-1.53	-1.22	-17.29	0.00
race(other)	-0.90	-1.11	-0.68	-8.15	0.00
education years	0.29	0.28	0.31	34.63	0.00
pack(<1)	0.25	0.11	0.39	3.46	0.00
pack($\geqslant$ 1)	0.37	0.26	0.48	6.73	0.00
diabetes	0.13	-0.03	0.29	1.62	0.10
hypertension	0.04	-0.06	0.15	0.80	0.43
high cholesterol	-0.08	-0.18	0.02	-1.56	0.12
trauma	0.06	-0.10	0.22	0.79	0.43

Table 6: Regression estimates on LMT score over time

	Est	ci95.lo	ci95.hi	z value	Pr(>|z|)
(Intercept)	27.56	19.99	35.12	7.14	0.00
year	1.15	0.86	1.45	7.77	0.00
dx(MCI)	31.29	29.60	32.99	36.16	0.00
dx(AD)	78.02	75.93	80.11	73.16	0.00
year*dx(MCI)	1.41	0.93	1.88	5.77	0.00
year*dx(AD)	5.22	4.63	5.82	17.23	0.00
age	2.01	1.93	2.10	49.08	0.00
female	-2.60	-4.19	-1.02	-3.22	0.00
race(AA)	37.43	35.02	39.84	30.43	0.00
race(other)	15.20	11.86	18.54	8.92	0.00
education years	-5.09	-5.34	-4.83	-39.03	0.00
pack(<1)	0.03	-2.16	2.22	0.03	0.98
pack($\geqslant$ 1)	-4.06	-5.75	-2.38	-4.73	0.00
diabetes	7.80	5.32	10.27	6.18	0.00
hypertension	2.12	0.51	3.74	2.57	0.01
high cholesterol	-1.89	-3.45	-0.33	-2.37	0.02
trauma	-1.60	-4.07	0.87	-1.27	0.20

Table 7: Regression estimates on TMT score over time

	Est	ci95.lo	ci95.hi	z value	Pr(>|z|)
(Intercept)	29.65	29.07	30.23	99.94	0.00
year	0.01	-0.00	0.03	1.41	0.16
dx(MCI)	-1.07	-1.16	-0.97	-21.29	0.00
dx(AD)	-2.55	-2.68	-2.42	-38.52	0.00
year*dx(MCI)	-0.06	-0.09	-0.04	-4.56	0.00
year*dx(AD)	-0.45	-0.48	-0.42	-25.72	0.00
age	-0.10	-0.11	-0.10	-32.06	0.00
female	-0.35	-0.48	-0.23	-5.66	0.00
race(AA)	-2.89	-3.08	-2.70	-30.27	0.00
race(other)	-1.82	-2.07	-1.56	-13.86	0.00
education years	0.29	0.27	0.31	28.71	0.00
pack(<1)	0.13	-0.04	0.30	1.53	0.13
pack($\geqslant$ 1)	0.41	0.28	0.55	6.23	0.00
diabetes	0.02	-0.17	0.21	0.18	0.85
hypertension	0.12	-0.01	0.24	1.83	0.07
high cholesterol	0.03	-0.09	0.16	0.56	0.57
trauma	0.25	0.06	0.44	2.54	0.01

Table 8: Regression estimates on BNT score over time

Test	LMT	TMT	BNT
Normal	0.43 (.00*)	-0.22 (.70)	0.20 (.00*)
MCI	0.47 (.00*)	1.74 (.08)	0.11 (.02*)
AD	-0.51 (.00*)	19.4 (.00*)	-1.13 (.00*)

Table 9: Short term summary

Test	LMT	TMT	BNT
Normal	0.24 (.00*)	1.15 (.00*)	0.01 (.16)
MCI	0.26 (.00*)	2.56 (.00*)	-0.05 (.00*)
AD	-0.09 (.09)	6.37 (.00*)	-0.44 (.00*)

Table 10: Longitudinal summary